DEVELOPING A STRATEGY FOR PARAMETER ESTIMATION FROM CO_2 PLUME MIGRATION DURING GEOLOGIC CARBON SEQUESTRATION IN A FLUVIAL DEPOSITIONAL SETTING

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Antoine Jean Espinet

January 2011

ABSTRACT

Estimation of the CO_2 plume from monitoring data is needed in order to assure safe carbon sequestration in geological formations. Synthetic field measurements and numerical simulations are used to estimate the plume position and obtain a better understanding of the characteristics of geological formations that govern the CO_2 flow. The challenge is to be able to give an accurate prognosis of the plume location with relatively few monitoring observations while dealing with uncertainties and model error. We use the TOUGH2 program, which is a numerical simulator for multi-phase fluid and heat flow in porous and fractured media, along with the ECO2N module, specific for CO_2 flow in brine. This model describes the coupling of flow and transport processes in heterogeneous geologic systems. The optimization program 'Stochastic RBF' is used to calibrate the model parameters. Stochastic RBF has proven to be computationally efficient for environmental models that are computationally expensive. It is a derivative-free method, which makes it easier to use in conjunction with a complex nonlinear simulation model. We use three-dimensional saline aquifers with different geological characteristics for the application. We show that estimating shale permeability is critical to determine the plume shape and position, while other facies higher permeabilities can be estimated with less accuracy with little effect on the estimate of the location of the CO_2 plume. We also investigate how parameter lumping affects the calibration and the amount of measurements needed in order to accurately estimate plume position. In many cases, it has been found that pressure measurements suffice, while other types of measurements are needed in cases with more parameters to estimate. The

plume position can be determined with a correlation coefficient equal to 87 percent with our method with a minimal amount of measurements and number of simulations. Using only pressure observations and no gas saturation samples, a slightly smaller correlation coefficient was obtained.

BIOGRAPHICAL SKETCH

Antoine Espinet was born in April 1987 in Reims, France. He received a Bachelor of Science in Fundamental Physics from the university of Orsay Paris Sud (France) in 2007 and in Engineering from the Ecole Centrale Paris in 2008. He graduated with a Master of Science in Engineering from the Ecole Centrale Paris in 2010. Antoine Espinet is currently a Ph.D candidate at Cornell University in the Civil and Environmental Department.

To my fiance and my parents

ACKNOWLEDGEMENTS

Thank you to my advisor Prof. Christine Shoemaker for her guidance, her time and financial support.

Thank you to Dr. Christine Doughty from the Lawrence Berkeley National Laboratory, without which this work could not have been carried out.

Thanks to Tianfu Xu, Karsten Pruess and Stefan Finsterle from the LBNL for their technical support and the Atkinson Center for a Sustainable Future (ACSF) and NSF grants (CBET 075675 and EAR 0711491) for their financial support to C. Shoemaker for A. Espinet. C. Doughty acknowledges financial support from the Assistant Secretary for Fossil Energy, Office of Sequestration, Hydrogen, and Clean Coal Fuels, National Energy Technology Laboratory, and by Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Geological carbon sequestration (GCS) has been proposed as a means to reduce greenhouse gas emissions. The most widely available sites are saline aquifers. When carbon dioxide is injected in a saline aquifer, its density can be lower than the surrounding fluid depending on the pressure and temperature conditions. The supercritical fluid is injected 1000 or more meters deep underneath a 'caprock' of very low permeability. As the injection goes on, the plume may reach some fracture in the caprock or abandoned wells with faulty seals, rising up and ruining the investment made in the GCS facility, and potentially carrying chemicals into fresh water resources. Industrial scale carbon sequestration will generate a plume that will extend over a surface area covering many square kilometers, making it very difficult and costly to monitor for leaks. For this reason, it is crucial to know where the CO_2 plume is located in order to focus our attention on weak areas (like abandoned wells or known fault lines) where we estimate that the CO_2 plume has reached. In addition, monitoring wells that are thousands of feet deep are expensive to construct, so plume estimation is going to be based on data from relatively few sampling points.

Saripalli and McGrail,2001 and *Nordbotten et al., 2005* have worked on estimation of the CO_2 plume position in homogeneous geological formations using analytical or semi-analytical solutions. However, these solutions cannot be applied to heterogeneous formations. *Weir et al., 1995* and *Pruess and Garcia, 2002* have carried out numerical simulations in homogeneous formations and *Doughty and Pruess, 2004*, *Juanes et al, 2006*, *Obi and Blunt, 2006*, and *Flett et al., 2007* and others numerically simulated three dimensional heterogeneous for-

mations. The focus of these papers is on processes of CO_2 storage and impact of various characteristics of the geological formation on CO_2 sequestration capabilities,. These detailed numerical models are called process-based. However in each cases, the geology is an input and is always treated as known. In real case however, the geology (e.g the spatial distribution and permeabilities of various facies) is not known precisely.

This paper focuses on estimating CO_2 plume and more precisely CO_2 gas saturation distribution based on limited monitoring data with use of a process-based model. This involves using numerical simulations optimization to solve the inverse problem, i.e. to characterize the unknown geology. Then the calibrated model is simulated forward in time with appropriate parameters to predict the CO_2 plume. We are focused on methods that are computationally efficient for processed-based simulation models. Model calibrations for geological carbon sequestration models have been carried out in *Bickle, 2007*, but the model was analytical (i.e not numerical simulations) and specific to the studied formation. *Doughty at al.,2008* also uses seismic data, as well as fluid samples and uses trial and error to best match field measurements. This only allows a small number of parameters to be estimated as the range for permeability for example, can extend over several orders of magnitude. Our optimization algorithm makes the estimation process fully automatic and more efficient in terms of both human time and computer time. The process automatic calibration van be repeated many times for updating as new monitoring data become available.

In this study, we will pose (1) the problem formulation, which includes describing the geological model used, the numerical simulator, as well as basic equations needed to carry out a calibration. Then, we will explain (2) our

methodology to tackle the problem, which includes the choice and setup of observation data, the choice of parameters and optimization algorithm, and finally the setup of the calibration problem and method to measure the goodness of the results. Finally, we will present (3) our results and discuss and interpret them. In particular, we will show how we manage a successful calibration with little data and a relatively small number of computationally expensive simulations, that is that we can determine the CO_2 plume position with over 80 percent accuracy.

In summary, this is the first study to do optimization-based calibration and CO_2 plume estimation for a detailed process-based numerical model of Geologic Carbon Sequestration. Explored is the value of different kinds of data (pressure versus gas saturation), the sensitivity of permeability parameters for different facies (body of rock with specified characteristics), location of sampling points and suitable ways to lump parameters. Because monitoring wells are very expensive, there will a limited number of monitoring points and we also evaluate the performance relative to the number of monitoring points.

CHAPTER 2

FORMULATION

This section presents the three-dimensional geological model and define the mathematical formulation used in the calibration.

2.1 Model Setup

The geological formation is inspired from the Frio CO_2 sequestration pilot site at the South Liberty oil field operated by Texas American Resources in Dayton, Texas, USA as described in *Hovorka et al., 2006*. There, 1,600 metric tons of CO_2 were injected over a period of 10 days into a steeply dipping brine-saturated sand layer at a depth of 1,500 m. At this depth, free phase CO_2 is supercritical (*Doughty et al., 2008*).

Our model is a generic model representing the fluvial/deltaic depositional environment of the Frio Formation in the upper Texas gulf coast (*Galloway, 1982; Hovorka et al., 2001*). The basic building blocks (the layers) of the model are generated stochastically, then the layers are superimposed to represent typical depositional sequences, in this case, a fining-upward storage formation, i.e. rocks with larger permeabilities at the bottom of the formation and with lower permeabilities at the top. The stochastic layer generation begins with schematic representations of three fluvial depositional settings found in this part of the Frio formation: barrier bars (continuous very high-permeability sands), distributary channels (intermingled sands and shales, with a large high-permeability sand component), and interdistributary bayfill (predominantly low-permeability discontinuous shale lenses, interspersed with moderate-permeability sand). The

program TProGS (*Carle and Fogg, 1996, 1997; Fogg et al., 2001*) uses transition probability theory to construct multiple two-dimensional stochastic representations of each depositional setting consistent with its schematic representation. Each realization honors the proportions, characteristic lengths, and juxtapositions of facies present in the schematic representation. More details on the stochastic generation process may be found in *Doughty and Pruess, 2004*.

the simulation assumes that CO_2 is injected for seven and a half years through a single injection well open over the lower half of the formation (i.e the CO_2 is injected only in the lower half of the formation). In order to draw meaningful conclusions, we created three stochastically generated formations, all upward fining and with the same properties (see Table 2.1). The three realizations (A, B and C) are made of 8000 rectangular grid blocks and 22,000 connections (see Figures 2.1, 2.2, 2.3).

Table 2.1: Geological Reservoir Properties

Dimensions	1 km*1 km for the fine grid and 100 meters thickness
Initial pressure	Hydrostatic, average: 150 bars
Initial Salinity	0.03 mass fraction of salt in brine
Temperature	60°C
Injection Rate of CO_2	$100,000$ tons per year or 3.16 kg/s

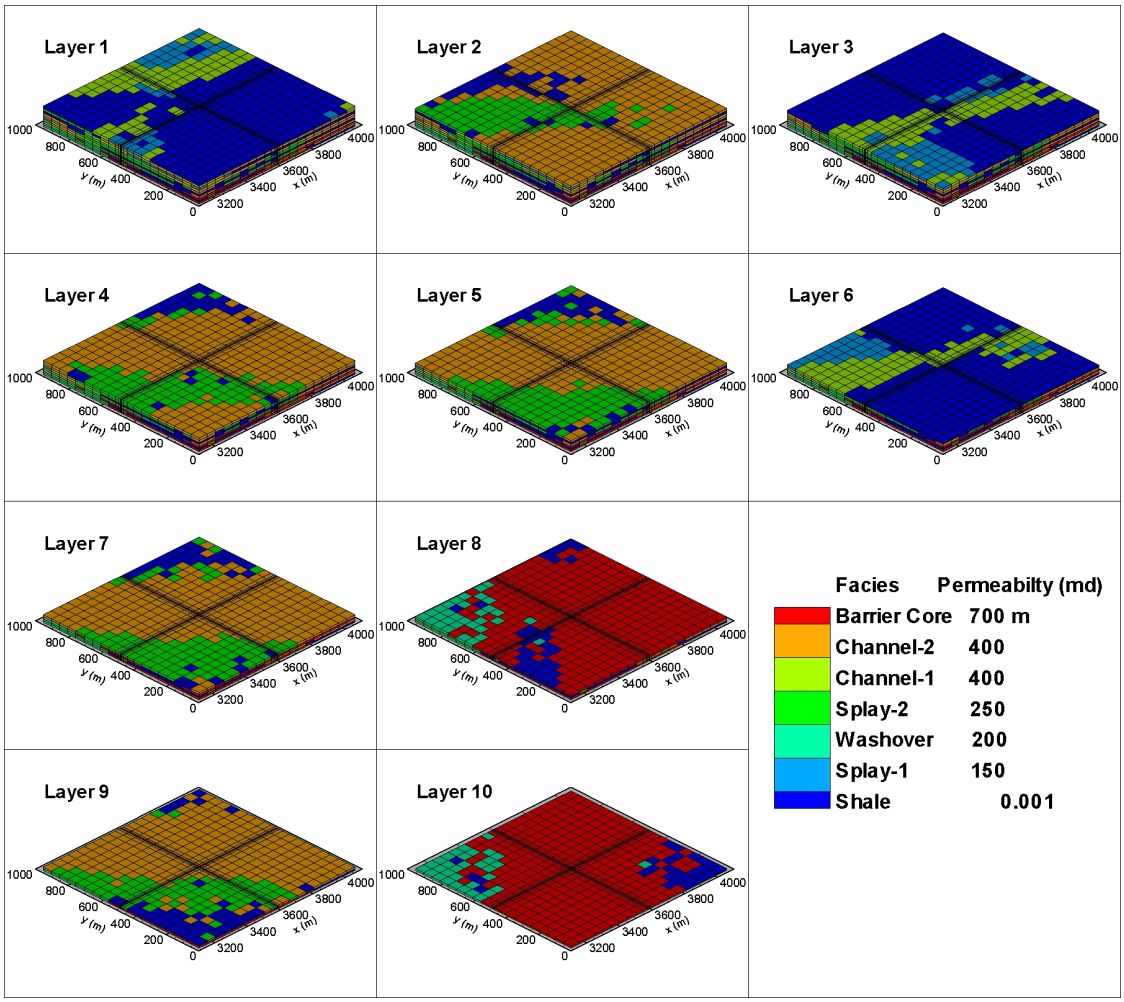

Figure 2.1: Three dimensional realization A, each color corresponds to a different facies

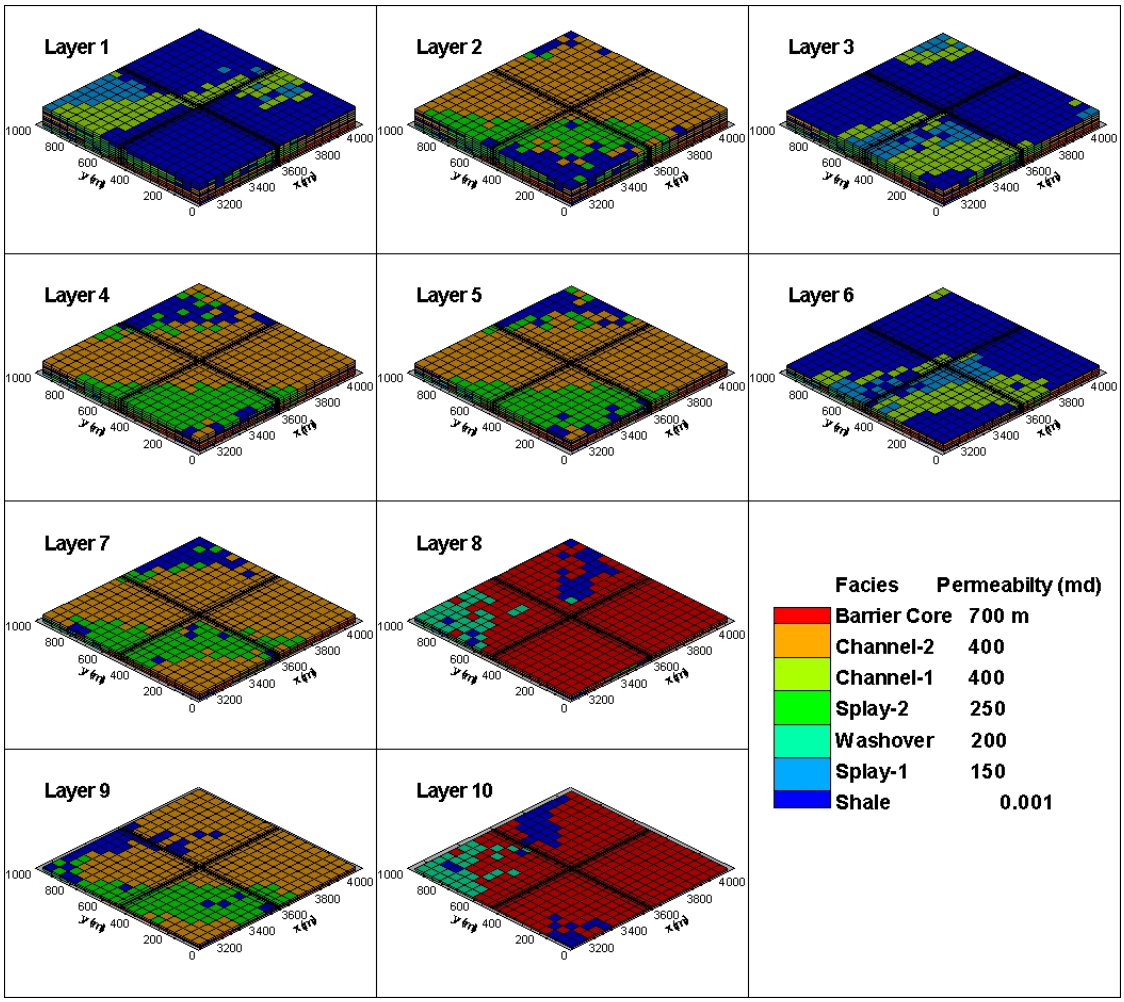

Figure 2.2: Three dimensional realization B, each color corresponds to a different facies

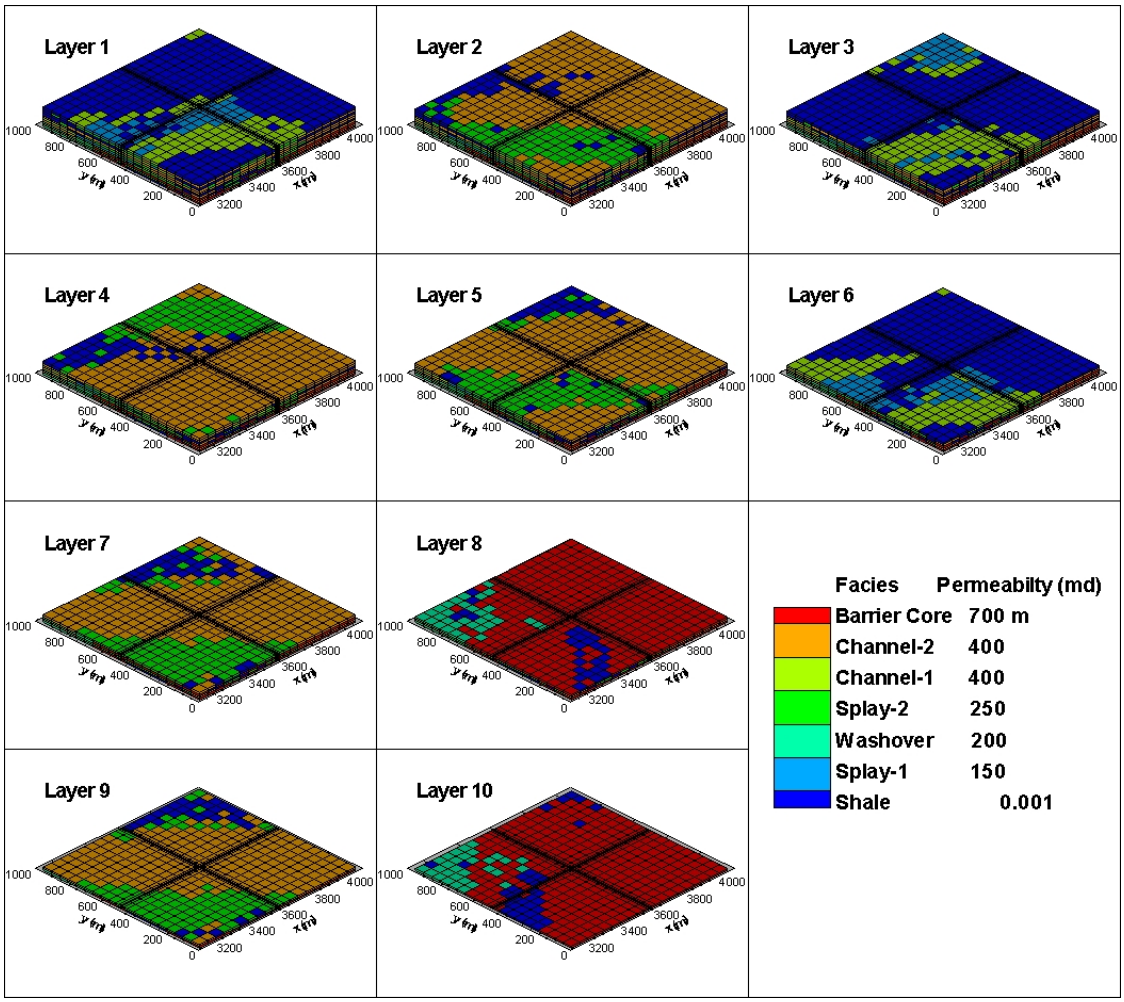

Figure 2.3: Three dimensional realization C, each color corresponds to a different facies

In order to simulate the CO_2 plume, we use the simulator TOUGH2 (*Pruess, 2004*) and the module ECO2N (*Pruess and Spycher, 2007*). TOUGH2 is a numerical simulation program for non-isothermal flows of multiphase, multicomponent fluids in permeable (porous or fractured) media. ECO2N is a fluid property module for the TOUGH2 simulator (Version 2.0) that was designed for applications involving geologic storage of CO_2 in saline aquifers. It includes a comprehensive description of the thermodynamics and thermo-physical prop-

erties of $H_2$2O-NaCl-CO_2 mixtures, that reproduces fluid properties largely within experimental error for the temperature, pressure and salinity conditions of interest. *Pruess et al., 2004* have shown that results on test problems using the TOUGH2 code and other CO_2 sequestration simulation codes have a high level of agreement, reducing the possibility of programming errors and numerical instabilities. The governing equations of multiphase and multicomponent flow are summarized in Appendix A.

2.2 Objective Function

The objective function is a classical sum-of-squares formulation, where we take the sum of the squares of the difference between the observed data and the simulated data at different times and locations. Hence, we can write:

$$Obj = \sum_{\{t,l,s\}\in\{T,L,S\}} \left(\frac{meas\,(t,l,s) - sim\,(t,l,s,p)}{\sigma_s} \right)^2 \tag{2.1}$$

where T - is the set of all times at which measurements are taken, L - is the set of all locations at which measurements are taken, S - is the set of all quantities that are measured (pressure, gas saturation...), $meas(\cdot)$ - is the measurement of species s that is observed at time t and location l, $sim(\cdot)$ - is the simulated data of species s at time t and location l, generated with the input parameter set p, σ_s - is the measurement error of the quantity s. We explain our choice of σ_s in the following section. p - is the parameter set that we are trying to calibrate. For example, $p = (p_1, ..., p_7)$, where p_i is the permeability of facies i (facies is represented by a color on Figures 2.1,2.2,2.3).

CHAPTER 3

METHODOLOGY

In this section , we will explain the generation of the observed data, the choice of locations for it and the innovative optimization algorithm that allows minimizing the objective function.

3.1 Observation data

Since no data are available for this synthetic site, synthetic observations of the pressure were made by running TOUGH2 with a given set of permeabilities p_0. Because the data is synthetic, we can modify our objective function to:

$$Obj = \sum_{\{t,l,s\}\in\{T,L,S\}} \left(\frac{sim\,(t,l,s,p_0) + err\,(s) - sim\,(t,l,s,p)}{\sigma_s} \right)^2 \tag{3.1}$$

where *err* is a zero mean normally distributed measurement error term.

A normal zero-mean random noise with standard deviation 7000 Pascals was added to the synthetic pressure data, to mimic measurement error. This represents an error mostly comprised between +/- 20,000Pa. The observations take place at the bottom of the injection well and at one or two observation wells. The number of observation wells is limited to two because they are very expensive to build and maintain. During the calibration period of 1.5 years, 100 samples are drawn from the two or three observation locations (more sampling during the transition phase and less when a near steady regime is reached). This frequency allows a good definition of the pressure response, meaning that the pressure response curve looks smooth. In practice, pressure samples could

be drawn much more frequently, but numerous trials showed that the higher-frequency sampling did not yield better results.

We also investigated the effect of observing the gas saturation in addition to the pressure. As for pressure, synthetic data was used. In the field, the gas saturation is not measured directly but instead a fluid sample is taken to the surface where its density ρ_{samp} is measured. The gas saturation, noted SG, is then deduced by applying the following formula:

$$\rho_{samp} = SG * \rho_{CO_2} + (1 - SG) * \rho_{brine} \tag{3.2}$$

The uncertainty in SG comes from the fact that heat loss occurs during the sampling process, causing some of the CO_2 to bubble out of the brine, hence changing SG slightly. Since we were not able to find a satisfying study on uncertainty on gas saturation measurement, we choose to add a zero mean normal noise with standard deviation 0.01, which gives a noise to signal ratio of the same order as the one used for the pressure observation.

3.2 Choice of the parameters

A preliminary study was carried out in order to establish which parameters were the most critical to determine the plume position. A Monte Carlo experiment was set up for a simpler homogeneous model with the following parameters: the permeability (horizontal and vertical permeability assumed to be equal), porosity, injection rate and the irreducible gas saturation for the van Genuchten-Mualem relative permeability function (*van Genuchten, 1980; Mualem, 1976*). Each parameter was given a uniform distribution that spanned

a feasible interval of values for the parameters. We repeatedly drew a value of each parameters randomly on its interval, simulated the CO_2 injection process, and computed the value of the objective function (using pressure data). This gave us a set of values of the objective function $Y = \{y_1, y_2, \ldots, y_n\}$, for which we can compute a variance $Var(y)$. Then we repeated this procedure, except that we kept one of the parameters x_i constant. This gives us a set of values of the objective function $CY^i = \left\{y_1^i, y_2^i, \ldots, y_n^i\right\}$, for which we can compute a variance $Var(y^i)$. Then the sensitivity $Sens(i)$ of the objective function due to the parameter x_i is defined as follow:

$$Sens\,(i) = Var\,(Y) - Var\left(CY^i\right) \tag{3.3}$$

Injection rate had a smaller sensitivity due to the fact that the chosen range of possible values was much smaller (relatively) in comparison to the permeability, which can cover several orders of magnitude, while the amount of CO_2 injected is well defined in practice, hence a smaller range of possible values. The permeability produced a much greater variance $Var(i)$ in the objective function than did the other parameters, and hence was chosen to be calibrated in this study. This result is confirmed for homogeneous formations in other studies like *Law and Bachu, 1996* and *Lindeberg, 1997*.

In Case 1, the parameters we are trying to calibrate are the permeabilities (horizontal and vertical separately) of six of the rock types in the formation as shown in Figures 2.1, 2.2, 2.3 (we suppose that the shale has a known permeability at first). The range of the permeabilities is [0.0001mD; 800mD]. There are 12 permeabilities to calibrate in total.

In Case 2, we do not assume that the permeability of the shale is known, which brings the number of parameters to calibrate to 14. However, we will show later that this number can be reduced significantly.

We do make some use of well information. We assume that we know the permeabilities of the rocks in the direct vicinity of the injection well. This is to prevent the optimization algorithm to propose a very low permeability value for theses rocks, which would mean injecting into a very low-permeability material, which would be numerically difficult. However, we assume that we don't know the properties of the rock around any observation well that we might use. This information is obtained during the well drilling.

3.3 The Stochastic RBF algorithm

Our objective function given by Equation 3.1 cannot be minimized by conventional optimization methods (like the derivative based Newton-Raphson or Levenberg-Marquart methods, or sequential quadratic programming or heuristic global optimization methods like Genetic algorithm) for the following reasons:

- The derivatives are not available or accurate because of computational expense and possible numerical error respectively hence, derivative based methods like Newton-Raphson or Levenberg-Marquardt are not usable
- The function has multiple local minima, hence local optimization algorithms do not ensure finding the best solution
- Most importantly, a simulation with our three dimensional model takes

up to 2 hours, hence the optimization algorithm should not need as many function evaluations as requested by heuristic optimization methods

A new algorithm presented in *Regis and Shoemaker, 2007* is used for optimization. It is tailored for expensive functions and does not use derivative information. It also handles global problems by using a multi-start method. A response surface (RBF: Radial Basis Function) is updated each time an simulation is made, giving an ever-improving approximation of the objective function curve. This way, no information is lost, reducing the number of simulation needed. The Stochastic RBF algorithm was previously applied to a complex nonlinear bioremediation groundwater model (*Mugunthan et al., 2005*) that took 3 hours per simulation. In this application, the Regis-Shoemaker method was compared to a number of other optimization algorithms and it was much more efficient. We also used the recently developed DDS (*Tolson and Shoemaker, 2007*) optimization algorithm because we had created a parallel version of it and could get answers in less wall clock time. Like the Stochastic RBF algorithm, it is tailored for expensive functions with multiple local minima and where derivatives aren't available or accurate. It has been proven efficient in a number of environmental problems.

3.4 Optimization-Calibration

In order to assess the accuracy with which we determined the plume position, we plot the observed gas saturation for all grid blocks versus the gas saturation for all grid blocks generated after running TOUGH2 with the calibrated set of parameters (called calibrated gas saturation for convenience). Then, we remove

the data points for which both the observed and calibrated gas saturation are zero (otherwise the correlation coefficient would be overinflated by the fact that a lot of points would have the coordinate (0,0)). Figure 3.1 shows an example of such a plot.

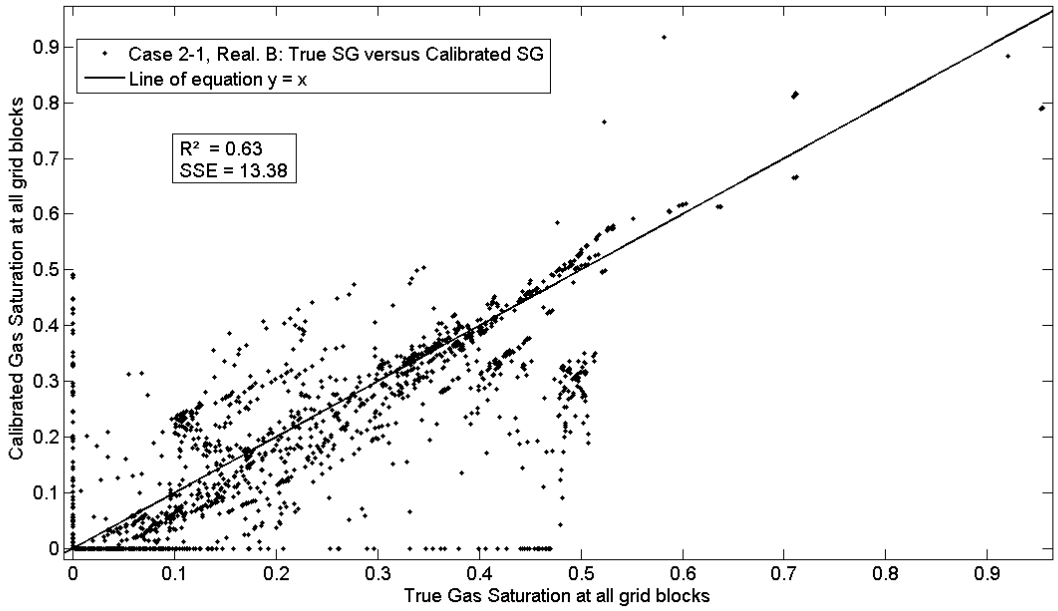

Figure 3.1: Plot showing how the calibration performance is assessed

Finally, the model is constrained so that line to obey the equation $y = x$. The correlation coefficient and sum of squared errors are computed and reported. In an ideal case, all points should be located on the line of equation $y = x$, i.e that the gas saturation distribution resulting from the calibrated parameters matches the true gas saturation distribution exactly.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Case 1

Case 1 is a simple case used for sensitivity analysis of parameters and monitoring locations.

4.1.1 Case 1 Setup

The shale permeability is assumed to be is known and set to its true value (known since it was used to generate the observation data). The vertical and horizontal permeability of all the other facies are calibrated separately (splay 1, splay 2, channel 1, channel 2, barrier core, and washover as shown in Figures 2.1, 2.2, 2.3), hence there are 12 parameters to calibrate. In Case 1, we run the calibration four times, each time with a different observation location (far from the shale, just underneath the shale, in between two shale layers and in the shale), in order to see the effect of the observation well location. Pressure information is used as the observation data. In all four calibrations, we limit ourselves to 150 evaluations, representing approximately 11 days of computer time per calibration.

4.1.2 Case 1 Results

The calibration did not succeed for any of the observation well locations, in the sense that the parameters were not at all close to their true value (although the plume was predicted with $R^2 \geq 0.60$). This meant either of three things: (1) the number of parameters was too large (search space is too big) to allow the algorithm to get to the minimum of the objective function in the limited number of function evaluations or (2) the observation data did not contain enough information about the parameter, or (3) the objective function was not sensitive to the chosen parameter set, that is that changes in the parameter set do not produce significant changes in the objective function (flat objective function). Note that ideas (2) and (3) are related but idea (2) is a statement about the observation data (the observation data is not sufficient or not relevant), whereas idea (3) is a statement about the choice of parameters (the parameters are not relevant to the plume shape in their current configuration).

We ran the optimization again with more evaluation (up to 1600) without better results, proving that the number of parameters was not the issue, hence disproving (1).

We therefore decided to investigate idea (2). We observed the pressure profile in the formation, presented in the figures 4.1 and 4.2. This shows that the pressure profile is very similar in space and time, when the shale permeability is known, no matter the other parameters chosen for the simulation. In other words, the pressure response seems to propagates independently of the facies, hence there is a small sensitivity of the objective function to the parameter set. Pressure in fact offers only a limited amount of information about the permeability of individual facies, as shown in Figure 4.3, which illustrates the

relationship between pressure change, porosity and gas saturation SG for all the grid blocks in the model. For high-porosity materials (which are the high-permeability materials, $SG \geq 0$), pressure change is positively correlated with gas saturation, indicating that the pressure response does provide some information on plume location, as stated in *Zhou et al., 2008*. However, for the low-porosity (low-permeability) shales ($SG = 0$, large pressure increases can occur where no CO_2 is present. In order to counterbalance the lack of information from the pressure, we added gas saturation monitoring and multi-depth observations at the same locations chosen for pressure observations.

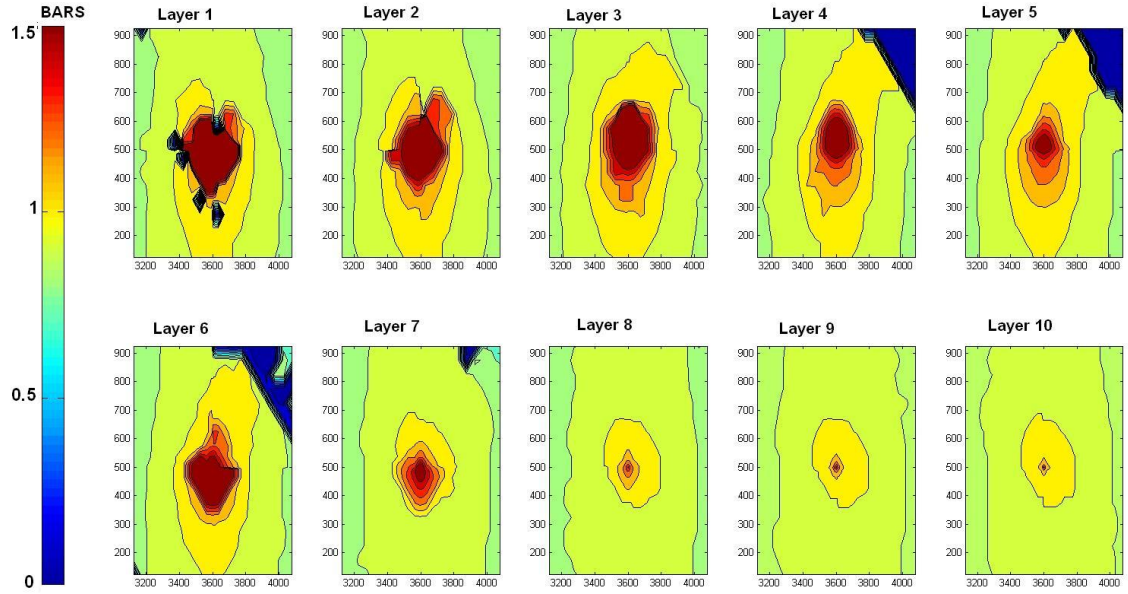

Figure 4.1: Pressure differential distribution at all depths for realization A at the end of the calibration period (1.5 years) generated with the true parameter values. Layer numbers refers to Figure 2.1

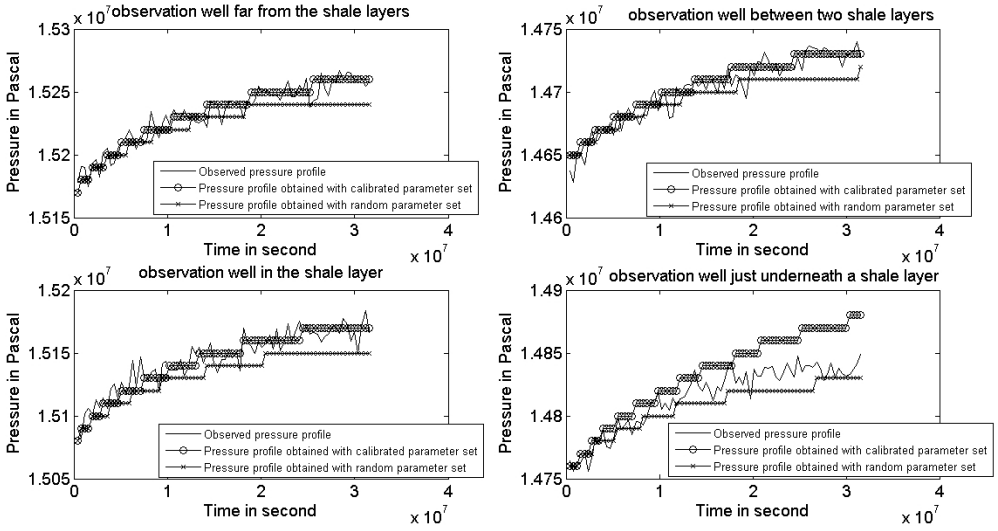

Figure 4.2: Pressure response at different locations for Case 1

Another important finding, supported by Figure 4.1 is that the location of an observation well that only monitors pressure is unlikely to matter since the pressure response is similar in a large vicinity of the injection site. However, as shown in Figure 4.4, the pressure response differs vertically, hence there will be a sensitivity of the calibration to the depth at which an observation well will be placed.

Idea (3) was also confirmed. We can see from Table 4.1 that the suspicion that the parameters were not relevant is confirmed. The plume position is determined with the same accuracy in all cases (i.e calibration with different observations wells as described above), including when we ran simulations with random parameter sets. This means that the parameter set chosen as described above does not allow the optimization algorithm to determine the CO_2 plume position. The parameters representing permeabilities of individual facies can be lumped together (see Case 2) to form a more sensitive and meaningful parame-

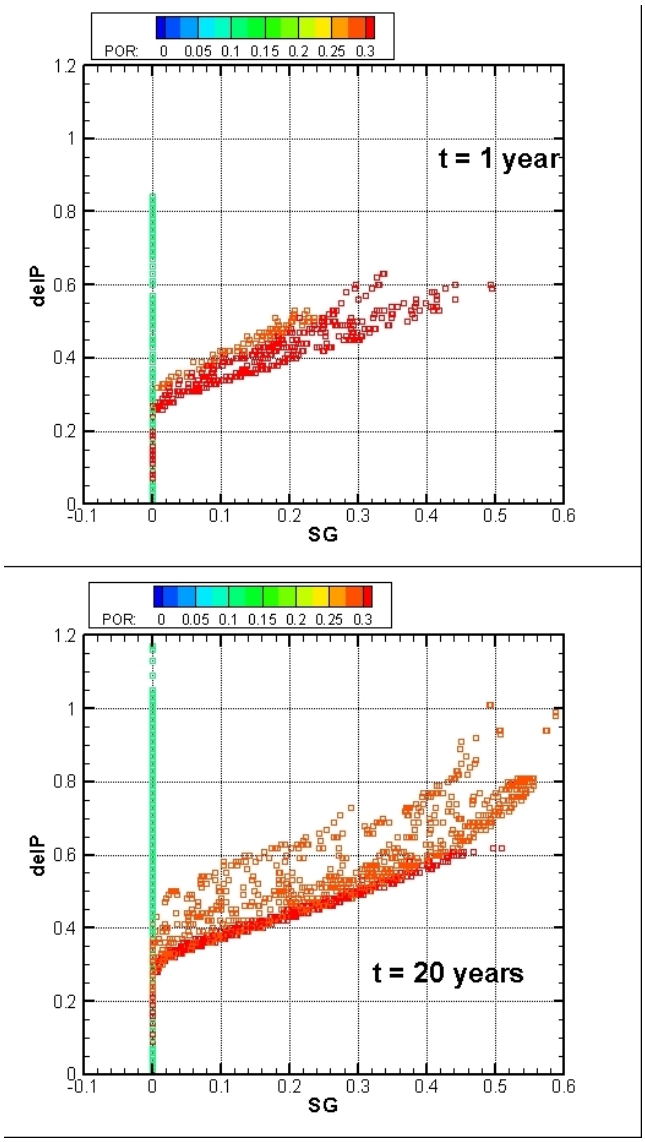

Figure 4.3: Pressure response versus gas saturation, generated with true parameter values

ter set.

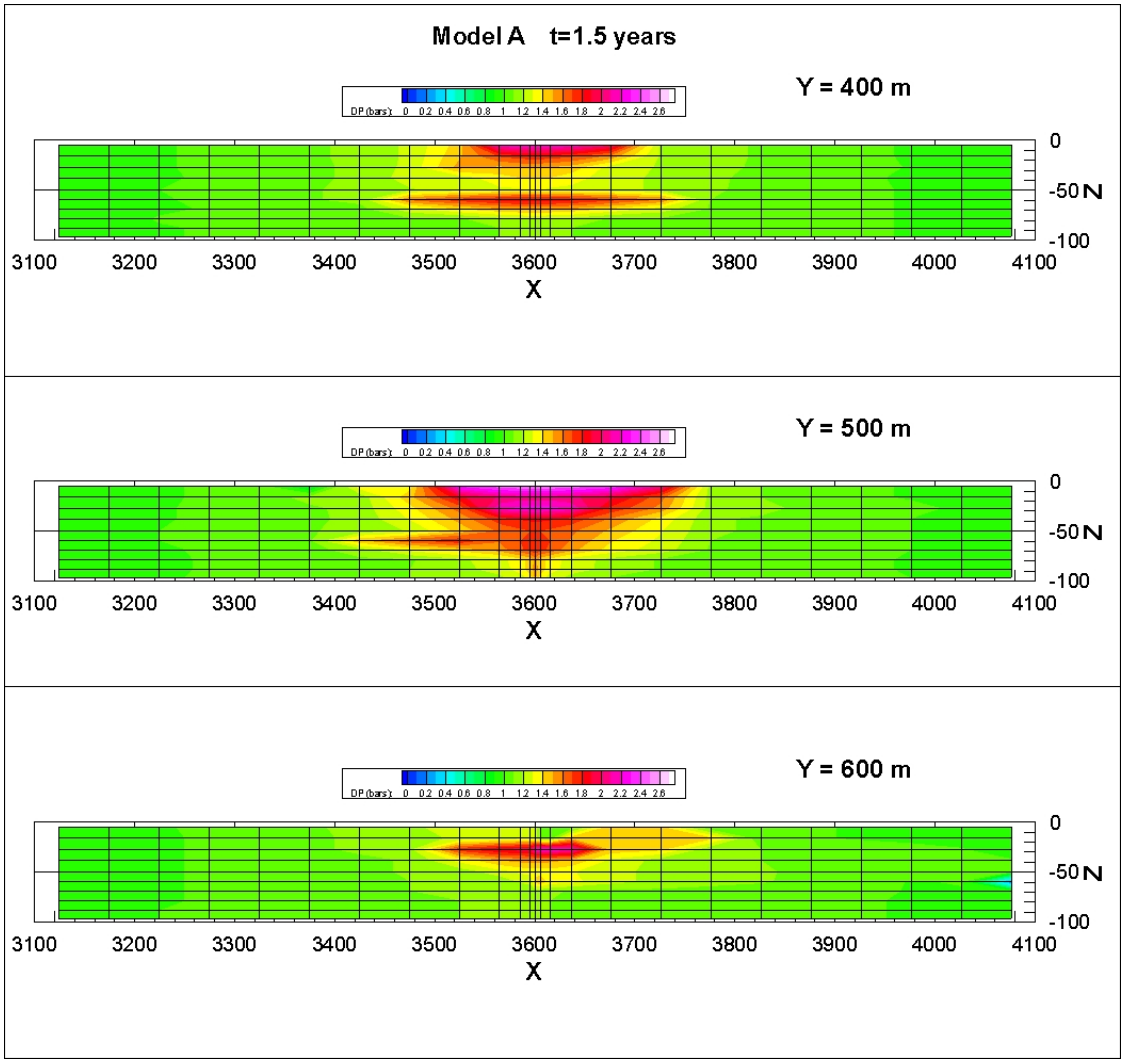

Figure 4.4: Pressure differential distribution for Case 1 and realization A at the end of the calibration period (1.5 years) generated with the true parameter values

Table 4.1: Goodness of fit obtained by running the realizations A, B and C with random parameter sets and also the one obtained in the calibration in Case 1

Case 1, scenario	Realization A	Realization B	Realization C
Observation far from shale	SSE = 7.23 R^2 = 0.72	SSE = 12.52 R^2 = 0.62	SSE = 10.24 R^2 = 0.63
Observation underneath the shale	SSE = 7.03 R^2 = 0.73	SSE = 12.89 R^2 = 0.60	SSE = 10.02 R^2 = 0.64
Observation in between the shale	SSE = 6.80 R^2 = 0.75	SSE = 12.63 R^2 = 0.61	SSE = 9.80 R^2 = 0.65
Observation in the shale	SSE = 7.03 R^2 = 0.73	SSE = 12.04 R^2 = 0.62	SSE = 10.22 R^2 = 0.65
random parameter set 1	SSE = 8.53 R^2 = 0.72	SSE = 13.21 R^2 = 0.63	SSE = 10.52 R^2 = 0.63
random parameter set 2	SSE = 8.2 R^2 = 0.74	SSE = 13.4 R^2 = 0.63	SSE = 10.89 R^2 = 0.61

Finally, Table 4.1 also shows that the plume position is determined with 60 to 75 percent accuracy, even when the parameter set is random (note that the shale permeability is known and fixed throughout the estimation process). In other words, it appears that knowing the shale permeability is the most critical aspect of plume estimation for carbon sequestration.

4.2 Case 2

Case 2 is a more complex, realistic description and it does result in better fits than Case 1.

4.2.1 Case 2 Setup

We do not assume that the shale permeability is known. This case is divided into 3 subcases: case 2-1 (Figure 4.5) has only two parameters: shale permeability and non-shale (called sand for convenience) permeability. Case 2-2 (Figure 4.6) has 3 parameters: the shale permeability, the gaps (rocks that form permeable pathways through the shale) and the permeability of the rest, called SAND for convenience. Case 2-3 (Figure 4.7) has 7 parameters: each shaly layer, i.e with more than 50 percent of shale in surface area, (layer 1, 3,and 6 so 3 in total) has its own shale and gap permeability (hence 6 parameters in total) and the permeabiltiy of the rest (called SAND for convenience). Note that Case 2 is closer to a real case scenario where the layers are identified (through seismic imaging for example) but of unknown permeability. It is also the case that when a well is drilled, the well might go through a shaly layer, giving information about at what depth shaly layers are located, but this process does not give any information about the extent of the shaly layer. In Case 2, we identify the shaly layers but also the extent of the shale as well as the pathways through the shale that are critical to the flow of CO_2 and hence the plume shape and position. In Case 2, we use the results from Case 1 and try different degrees of lumping as described above. In order to assess how much and what observed data is needed to successfully estimate the plume position, we ran multiple calibrations with

different data points (pressure only and pressure and gas saturation together), and with different numbers of monitoring locations (single depth, multi-depth, number of monitoring wells).

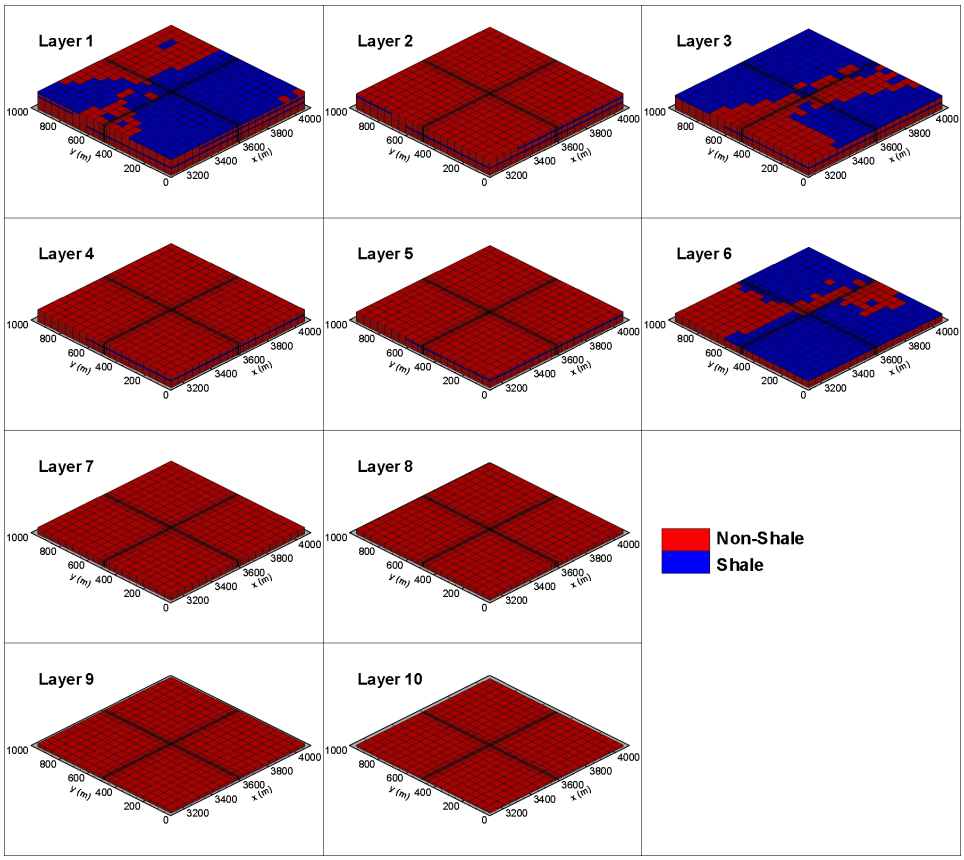

Figure 4.5: Case 2-1, 2 parameters: the shale and the sand permeabilities

[Only showed for realization A]

The observation well mostly used and noted 'OW' ('Observation Well') is 50 meters from the injection well. It was necessary to make it so close in order to see a gas saturation response by the end of the calibration period. For a greater amount of injection and a resulting larger plume, the monitoring well could be further away from the injections site. We also use a second observation well in cases 2-2 and 2-3, noted 'far OW', which is 400 meters from the injection well. This observation well is only used for pressure information, as the gas satura-

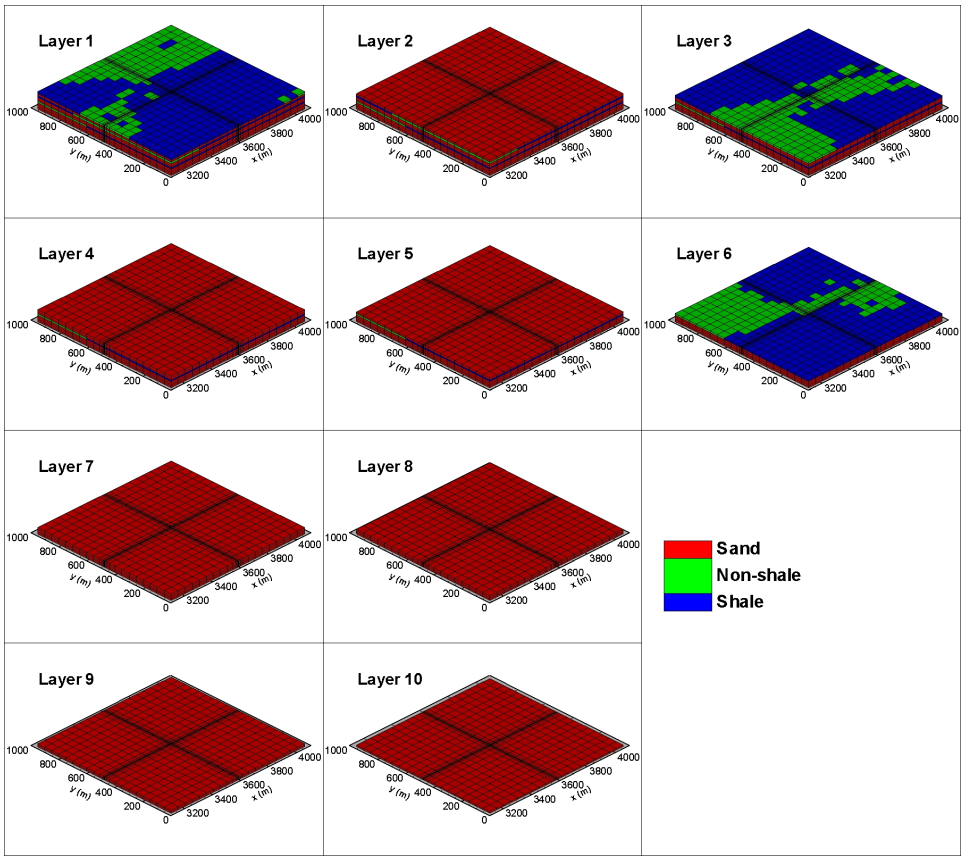

Figure 4.6: Case 2-2, 3 parameters: the shale, the gaps and the sand permeabilities

[Only showed for realization A]

tion at the end of the calibration period is still zero (the CO_2 has not yet reached that point) but the pressure response is positive since the pressure field can extend much further than the plume (i.e in the brine) as discussed in *Zhou et al., 2008*. As shown in Case 1, the location of the observations well is of little importance because the pressure response propagates throughout the formation, no matter the different permeability zones as shown in Case 1. However, in practice, it is known that the most effective depth to sample from is just below a shale layer because the pressure response will be most different from a parameter set to another at these locations as explained later in this paragraph. That is the reason why our samples are always drawn from layers 2, 4 and 7

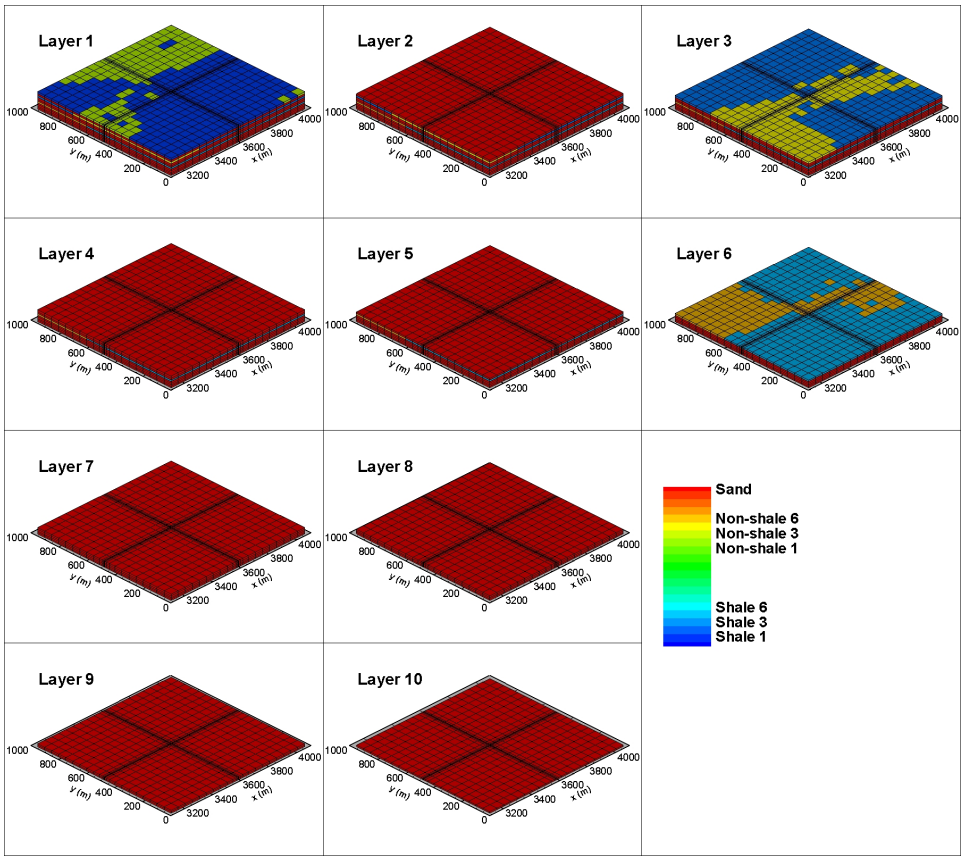

Figure 4.7: Case 2-3, 7 parameters: the shale and gaps permeabilities at layers 1, 3 and 6 and the sand permeability

[Only showed for realization A]

(see Figures 2.1, 2.2, and 2.3). Such a strategy allows the samples to carry more information about the shale permeability (which we showed to be critical in the plume position prediction process). In fact, the pressure response will be very different if the parameter chosen by the optimization algorithm for the shale permeability is large (say more than 10mD), allowing the CO_2 to go through the layer rather than if the parameter chosen by the optimization algorithm for the shale permeability is small (say less than 0.1mD), forcing the CO_2 to accumulate underneath it.

For all three subcases, we impose the number of function evaluations. Each

time, we plot the objective function versus the number of evaluations in order to make sure that a minimum was reached, i.e. that the objective function cannot be minimized further by additional evaluations. In practice, the minimum value of the objective function found by the optimization algorithm will drop quickly and then flatten out as the number of evaluations increases. This does not guarantee that the optimization algorithm has found the global minimum, but it makes it more probable. On average, it is worthwhile mentioning the algorithm found the reported minimum of the objective function in half the number of evaluations used, suggesting that no more simulations were needed.

4.2.2 Case 2 Results

Observe that the objective function is much less flat in Case 2 than in Case 1 thanks to Figures 4.8, 4.9, and 4.2. The value of the objective function (Figure 4.8) in Case 1 can only be improved by approximately 0.5 (the value of the vertical axis is the logarithm of the value of the objective function), while in Case 2, it is improved by over 9. Also, the pressure responses for case 2 (Figure 4.9) generated from a random set of parameters look much different than the ones generated from the true or calibrated parameter set as opposed to Case 1 (Figure 4.2). This proves again that including the shale permeability in the estimation makes the calibration much more difficult.

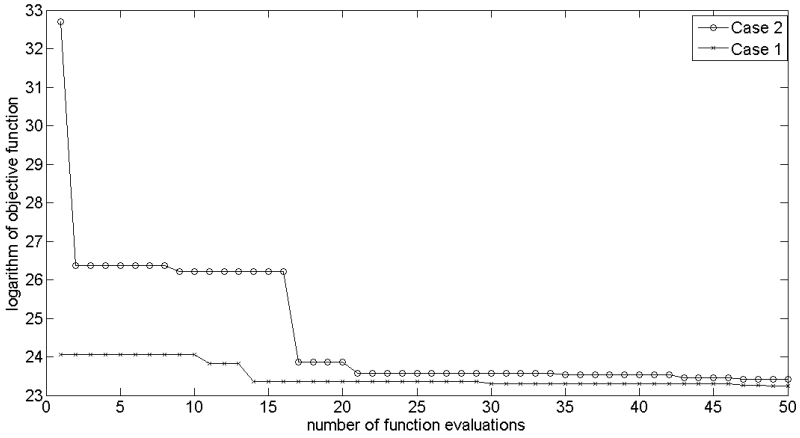

Figure 4.8: Objective function value for Cases 1 and 2 (here for Case 2-1) versus the number of evaluations during the calibration process

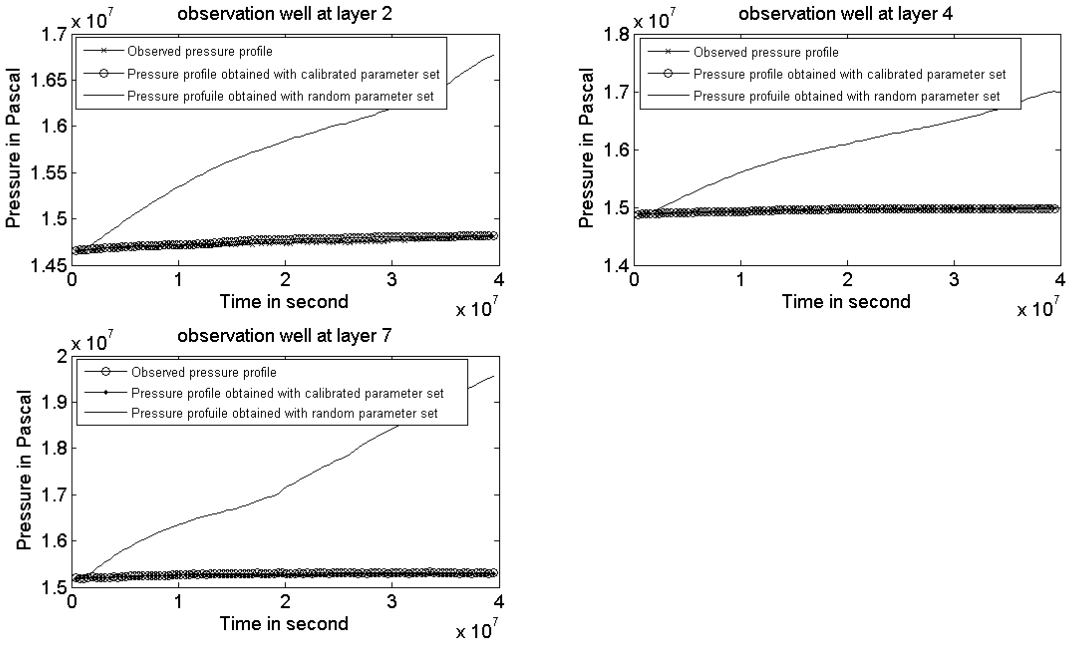

Figure 4.9: Pressure response at different locations for Case 2

The results for case 2-1, summarized in Table 4.2 confirm that knowing the permeability of the shale layer is very important in determining the position of the CO_2 plume. With only two parameters, the correlation coefficient was about 60 to 75 percent (R^2 reported using the calibration assessment technique described before). For two parameters, we notice that little data is needed: monitoring the pressure at two locations (bottom of the injection well and at an observation well 300 meters away just underneath the top shale layer) allows to correctly estimate the formation parameters, i.e that the shale layers is identified as such (low permeability, 0.01mD on average for the three realizations) and the rest of the geology is identified as a higher permeability region (300mD on average for the three realizations). Measurement of the gas saturation allows gaining only 2 percent for the correlation coefficient, which is negligible in relation to the cost of sampling the fluid.

In case 2-2, (Table 4.3), we observe that the plume position is revealed with more accuracy than in case 2-1 (the correlation coefficient representing the fit between the true plume and the estimated plume differs by only 2 percent for realization A and 10 percent for realization B and C). The difference with case 2-1 is that we separated the estimation of the permeability of the gaps in the shale layers from the rest. More data is needed for Case 2-2 than for Case 2-1 for the calibration and more function evaluation are performed, which is explained by the increased number of parameters to estimate. But again, pressure measurements alone suffice to allow the optimization algorithm a good solution. We can also conclude that adding gas saturation data only improves the fit slightly (2 percent on average), while costing a significant increase in cost for fluid sampling and analysis. The important conclusion here is that meaningful lumping of the parameters allow the objective function to be sensitive enough

Table 4.2: Calibrated parameter set and goodness of fit for Case 2-1

[P: Pressure, SG: Gas Saturation, IW: Injection well, OW: Observation well, L4: layer 4 as shown in Figure 2.1. Real: realization; Observed data: data points measured and included in Objective function as stated in equation 3.1; Number of TOUGH2 runs: number of simulations; Normalized Obj.: objective function as stated in equation 3.1 divided by the number of data points; R^2(SSE): correlation coefficient computed as explained in section Optimization-Calibration and corresponding Sum of Squared Errors]

Case 2-1	Real. A	Real. B	Real. C
Observed data	200 data points: P only at IW L10, OW L2		
Number of TOUGH2 runs	50		
Normalized Obj	1.67	3.70	3.02
R^2	0.73(7.79)	0.63(13.38)	0.64(10.86)
Observed data	400 data points: P, SG at IW L10, OW L2		
Number of TOUGH2 runs	50		
Normalized Obj.	1.51	3.57	2.34
R^2(SSE)	0.75(8.12)	0.64(11.85)	0.67(9.47)

for the optimization algorithm to find a good solution.

The case 2-3 (Table 4.4) is the most general case. It does not assume that the shale layers and gaps have the same permeabilities, which brings the number of parameters to 7 (3 shaly layers, each with gaps and the rest). The consequence is that more data and more function evaluations are needed. With 1400 data points

Table 4.3: Calibrated parameter set and goodness of fit for Case 2-2

P: Pressure, SG: Gas Saturation, IW: Injection well, OW: Observation well, L4: layer 4 as shown in Figure 2.1. Real: realization; Observed data: data points measured and included in Objective function as stated in equation 3.1; Number of TOUGH2 runs: number of simulations; Normalized Obj.: objective function as stated in equation 3.1 divided by the number of data points; R^2(SSE): correlation coefficient computed as explained in section Optimization-Calibration and corresponding Sum of Squared Errors

Case 2-2	Real. A	Real. B	Real. C
Observed data	400 data points total: P at IW L10, far OW L2,4,7		
Number of TOUGH2 runs	120		
Normalized Obj.	1.32	1.72	1.55
R^2(SSE)	0.78(6.21)	0.74(7.30)	0.75(7.71)
Observed data	800 data points total: P, SG at IW L10, OW L2,4,7		
Number of TOUGH2 runs	120		
Normalized Obj.	1.11	1.41	1.35
R^2 (SSE)	0.80(5.51)	0.77(6.32)	0.77(6.40)

and 200 function evaluations, the calibration exceeds the fit level obtained in Case 2-2, with up to a 10 percent gain in accuracy in the plume position. We can see that much more data is needed. In particular, an other observation well is needed for additional pressure measurement. With pressure data only, the fit is improved compared to Case 2-2, but not as much as with additional gas

saturation data. Figure 4.10 shows the plume at layer 4 for the true case, the calibrated case and a case where the parameters have been set arbitrarily. The vertical slice in figure 4.11 shows the agreement between the true and calibrated plumes.

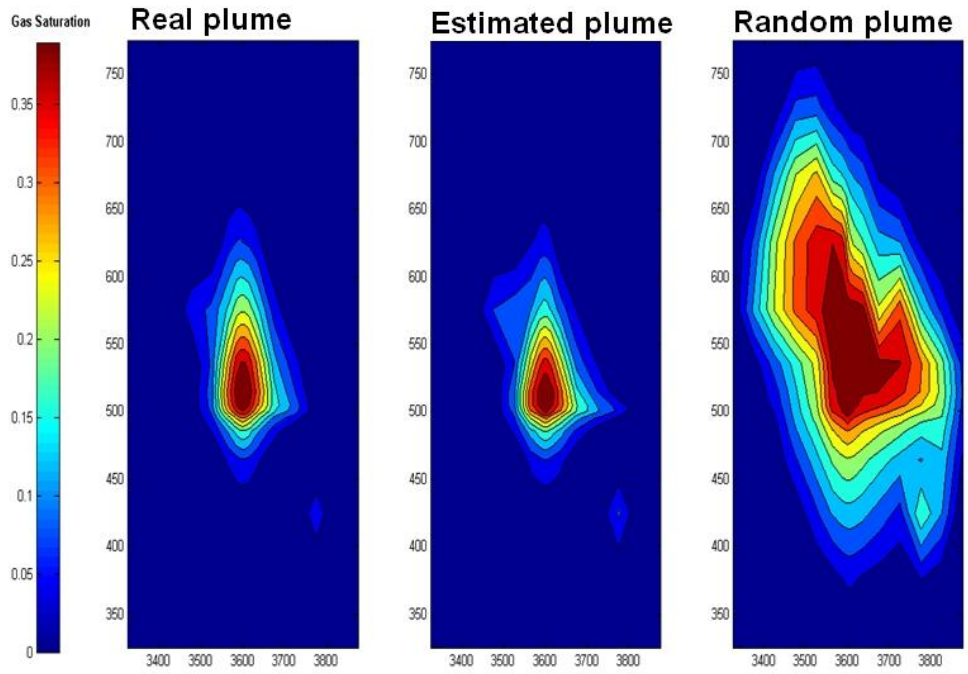

Figure 4.10: Plume at layer 4

[From left to right: true plume, plume resulting from calibrated parameters, and plume from randomly generated parameters for Case 2-3, realization C]

Table 4.4: Calibrated parameter set and goodness of fit for Case 2-3

P: Pressure, SG: Gas Saturation, IW: Injection well, OW: Observation well, L4: layer 4 as shown in Figure 2.1. Real: realization; Observed data: data points measured and included in Objective function as stated in equation 3.1; Number of TOUGH2 runs: number of simulations; Normalized Obj.: objective function as stated in equation 3.1 divided by the number of data points; R^2(SSE): correlation coefficient computed as explained in section Optimization-Calibration and corresponding Sum of Squared Errors

Case 2-3	Real. A	Real. B	Real. C
Observed data	700 data points: P at IW L10, OW L2,4,7 and at far OW L2,4,7		
Number of TOUGH2 runs	200		
Normalized Obj.	0.97	1.46	1.15
R^2 (SSE)	0.83(4.53)	0.77(6.26)	0.80(5.94)
Observed data	1100 data points total: P, SG at IW L10, OW L2,4,7 and P at far OW L2,4,7		
Number of TOUGH2 runs	200		
Normalized Obj.	0.87	1.10	0.99
R^2 (SSE)	0.87(3.82)	0.81(6.32)	0.87(6.33)

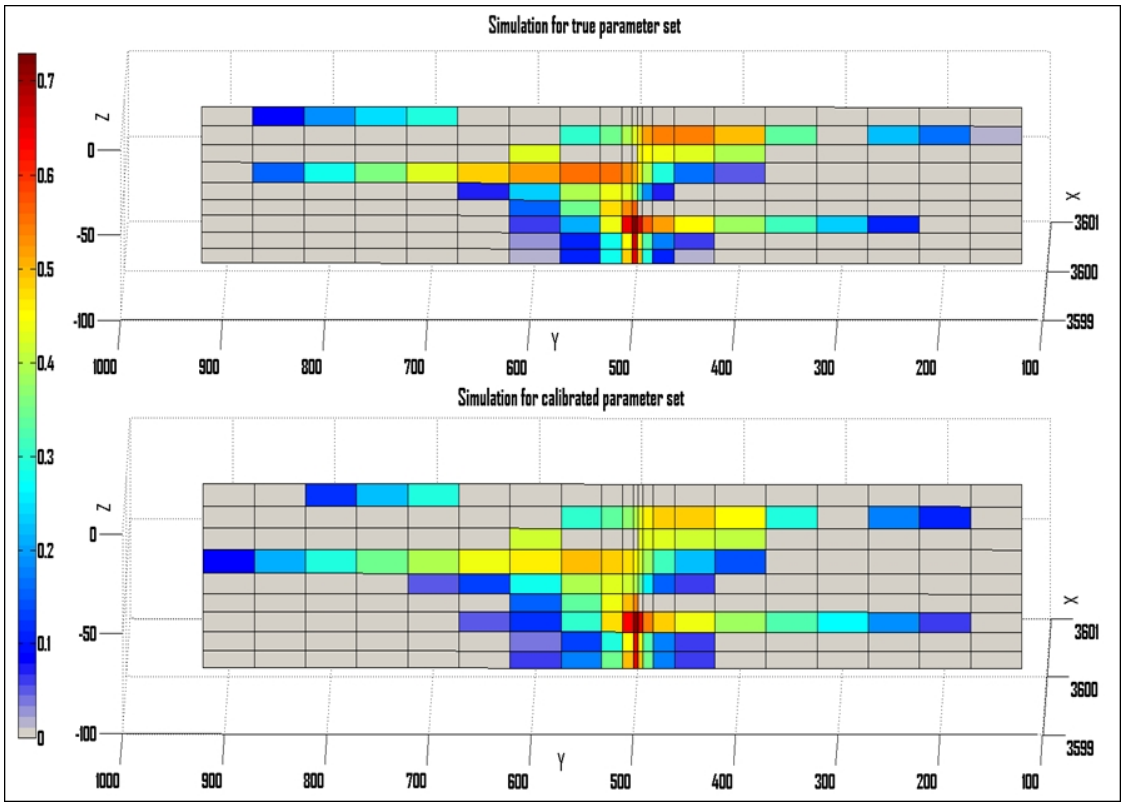

Figure 4.11: Vertical cut of formation, passing through the injection well

[Top: true plume. Bottom: plume generated from calibrated parameters for Case 2-3, realization C]

As shown in Figures 4.10 and 4.11, a correlation coefficient of 87 percent provides a good level of detail about the plume shape. In particular, preferential pathways are well identified on Figure 4.11. The surface area occupied by the plume is also well identified in Figure 4.10

CHAPTER 5

CONCLUSION

In this work, it has been shown firstly that the permeability of the shale layer is crucial to determining the plume position with any accuracy. Secondly, from multiple calibrations with different data, at different locations, we can conclude that the pressure information alone is sufficient to estimate correctly the plume position, at the condition that the parameters are the product of meaningful lumping. In particular, a simple lumping between shale and non-shale elements has proved to yield a good first approximation of the plume position. When pressure samples only are drawn, the locations of the wells is of little importance, but samples are chosen to be drawn just below shaly layers. Finally, a high correlation coefficient for the plume position was obtained (see Case 2-3), using only pressure data and two monitoring wells. It is worthwhile noting that this study was carried out for much smaller than industrial scale formations, but we believe that our conclusions still apply. In fact, larger injections rates for example, would generate a bigger signal to noise ratio, making it easier for the optimization algorithm to find the minimum of the objective function. Future work includes studying formations with an upward coarsening depositional setting, as well as formations with a low permeability barrier.

APPENDIX A

TOUGH2 SIMULATOR GOVERNING EQUATIONS

These equations and additional details on TOUGH2 are presented in *Pruess, 2004*. More details can be found in the TOUGH2 use's guide (*Pruess et al., 1999*).

We write the classical mass conservation equation for multiphase and multicomponent flow.

$$\frac{d}{dt}\int_{V_n} M^K dV_n = \int_{\Gamma_n} \mathbf{F^K}\cdot\mathbf{n}d\Gamma_n - \int_{V_n} q^K dV_n \tag{A.1}$$

where V_n is an arbitrary domain of study, bounded by the closed surface Γ_n. M^K represents the mass per volume of component K (brine or CO_2). $\mathbf{F}$ denotes the mass flux and q the sinks or sources. $\mathbf{n}$ is a normal vector on surface element $d\Gamma_n$, pointing outward.

We define M^K to be the sum over the mass all fluid phases of component K

$$M^K = \phi\sum_\beta S_\beta\rho_\beta X_\beta^K \tag{A.2}$$

where ϕ is the porosity, S_β the saturation of phase β (i.e the fraction of pore volume occupied by phase β), ρ_β the density of phase β, and X_β^K the mass fraction of component K in phase β.

The flux $\mathbf{F}$ has two components. One is an advective component:

$$F_{adv}^K = \sum_\beta X_\beta^K \mathbf{F}_\beta \tag{A.3}$$

where we define $\mathbf{F}_\beta$ as:

$$F_\beta = \rho_\beta \mathbf{u}_\beta = -k\frac{k_{r\beta}\rho_\beta}{\mu_\beta}\left(\nabla P_\beta - \rho_\beta \mathbf{g}\right) \tag{A.4}$$

where $\mathbf{u}_\beta$ is the Darcy velocity (volume flux) in phase β, k is the absolute permeability, $k_{r\beta}$ is the relative permeability to phase β, μ_β the viscosity, $\mathbf{g}$ the gravitational acceleration, and the fluid pressure in phase β P_β is defined as:

$$P_\beta = P + P_{c\beta} \tag{A.5}$$

where P is the pressure of a reference phase (usually the gas phase) and $P_{c\beta}$ the (negative) capillary pressure.

The other component of the flux is the diffusive flux of component K in phase β:

$$\mathbf{f}_\beta^K = -\phi\tau_0\tau_\beta\rho_\beta d_\beta^K \nabla X_\beta^K \tag{A.6}$$

where $\tau_0\tau_\beta$ is the tortuosity, which includes a porous medium dependent factor τ_0 and a phase saturation S_β dependent coefficient τ_β. d_β^K is the diffusion coefficient of component K in phase β.

BIBLIOGRAPHY

[1] Bickle, M; Chadwick, A; Huppert, HE; et al.: Modelling carbon dioxide accumulation at Sleipner: Implications for underground carbon storage, *Earth and Planetary Science Letters*, 255 (1-2): 164-176 MAR 15 2007

[2] Carle, SF; Fogg, GE: Transition probability-based indicator geostatistics, *Mathematical Geology*, 28 (4): 453-476 MAY 1996

[3] Carle, SF; Fogg, GE: Modeling spatial variability with one and multidimensional continuous-lag Markov chains, *Mathematical Geology*, 29 (7): 891-918 OCT 1997

[4] Doughty, C. and K. Pruess: Modeling supercritical carbon dioxide injection in heterogeneous porous media, *Vadose Zone Journal*, 3, 3, 837-847, 2004

[5] Doughty, C; Freifeld, BM; Trautz, RC: Site characterization for CO_2 geologic storage and vice versa: the Frio brine pilot, Texas, USA as a case study, *Environmental Geology*, 54 (8): 1635-1656 JUN 2008

[6] Flett, M; Gurton, R; Weir, G: Heterogeneous saline formations for carbon dioxide disposal: Impact of varying heterogeneity on containment and trapping, *Journal of Petroleum Science and Engineering*, 57 (1-2): 106-118 MAY 2007

[7] Fogg, G.E., S.F. Carle, and C.T. Green, A connected network paradigm for the alluvial aquifer system, in D. Zhang and C.L Winter, eds.,Theory, modeling, and field investigation in hydrogeology: a special volume in honor of Shlomo P. Neumans 60th birthday, *Geol. Soc. of Am.*, Special Paper 348, 2001

[8] Galloway, WE: Frio Formation of Texas Gulf Coastal-Plain - Depositional systems, structural framework, and hydrocarbon distribution, *AAPG Bulleting-American Association Of Petroleum Geologists*, 66 (6): 649-688 1982

[9] Hovorka, S.D., C. Doughty, P.R. Knox, C.T. Green, K. Pruess, and S.M. Benson, Evaluation of brine-bearing sands of the Frio formation, upper Texas gulf coast for geological sequestration of CO_2, *First National Conference on Carbon Sequestration*, May 14-17, Washington DC, National Energy Technology Laboratory, 2001

[10] Hovorka, S.D., S.M. Benson, C. Doughty, B.M. Freifeld, S. Sakurai, T.M. Daley, Y.K. Kharaka, M.H. Holtz, R.C. Trautz, H.S. Nance, L.R. Myer, and

K.G. Knauss: Measuring permanence of CO_2 storage in saline formations: the Frio experiment, *Environmental Geosciences*, 13, 2, 1-17, 2006

[11] Juanes, R; Spiteri, EJ; Orr, FM; et al.: Impact of relative permeability hysteresis on geological CO_2 storage, *Water Resource Research*, 42 (12): Art. No. W12418 DEC 23 2006

[12] Law, DHS; Bachu, S: Hydrogeological and numerical analysis of CO_2 disposal in deep aquifers in the Alberta Sedimentary Basin, *Energy Conservation and Management*, 37 (6-8): 1167-1174 JUN-AUG 1996

[13] Lindeberg, E: Escape of CO_2 from aquifers, *Energy Conservation and Management*, 38: S235-S240 Suppl. S 1997

[14] Mualem, Y., New model for predicting hydraulic conductivity of unsaturated porous-media, *Water Resource Research*, 12 , 513-522, 1976

[15] Mugunthan, P., Shoemaker, C. A., Regis, R. G: Comparison of function approximation, heuristic, and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models, *Water Resource Research*, 41 (11) : w11427 2005

[16] Nordbotten, JM; Celia, MA; Bachu, S: Injection and storage of CO_2 in deep saline aquifers: Analytical solution for CO_2 plume evolution during injection, *Transport in Porous Media*, 58 (3): 339-360 MAR 2005

[17] Pruess, K; Oldenburg, C; Moridis G: TOUGH2 USERS GUIDE, VERSION 2.0, Earth Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, 1999.

[18] Pruess, K; Garcia, J: Multiphase flow dynamics during CO_2 disposal into saline aquifers, *Environmental Geology*, 42 (2-3): 282-295 JUN 2002

[19] Pruess, K: The TOUGH codes - A family of simulation tools for multiphase flow and transport processes in permeable media, *Vadose Zone Journal*, 3 (3): 738-746 AUG 2004

[20] Pruess, K; Spycher, N: ECO2N - A fluid property module for the TOUGH2 code for studies of CO_2 storage in saline aquifers, *Energy Conservation and Management*, 48 (6): 1761-1767 JUN 2007

[21] Regis, RG; Shoemaker, CA: A stochastic radial basis function method for

the global optimization of expensive functions, *Informs Journal of Computing*, 19 (4): 497-509 FAL 2007

[22] Saripalli, P; McGrail, P: Semi-analytical approaches to modeling deep well injection of CO_2 for geological sequestration, *Energy Conservation and Management*, 43 (2): 185-198 JAN 2002

[23] Schnaar, G; Digiulio, DC: Computational Modeling of the Geologic Sequestration of Carbon Dioxide, *Vadose Zone Journal*, 8 (2): 389-403 MAY 2009

[24] Tolson, B. A., Shoemaker, C. A: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resource Research*, 43(1): W01413 2007

[25] Van Genuchten, M. T., A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, 44 , 892-898, 1980

[26] Zhou, QL; Birkholzer, JT; Tsang, CF; et al.: A method for quick assessment of CO_2 storage capacity in closed and semi-closed saline formations, *International Journal of Greenhouse Gas Control*, 2 (4): 626-639 Sp. Iss. SI OCT 2008